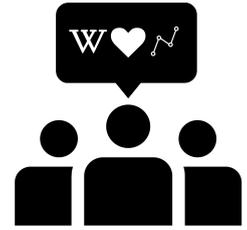


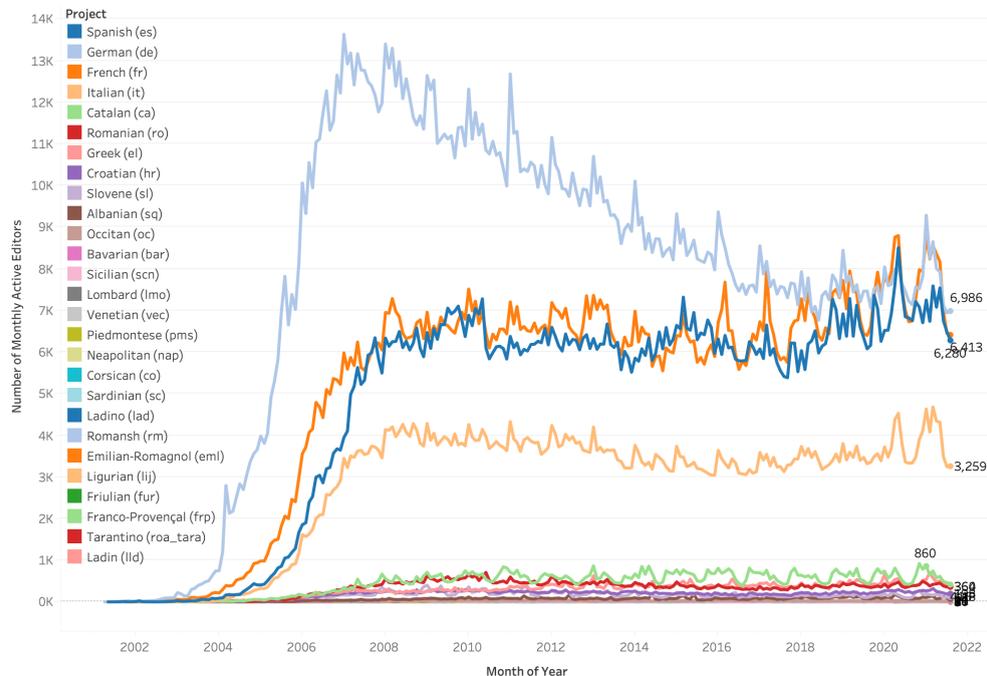
Italian Wikipedias Vital Signs Community Health Metrics

Marc Miquel, David Laniado, Cristian Consonni

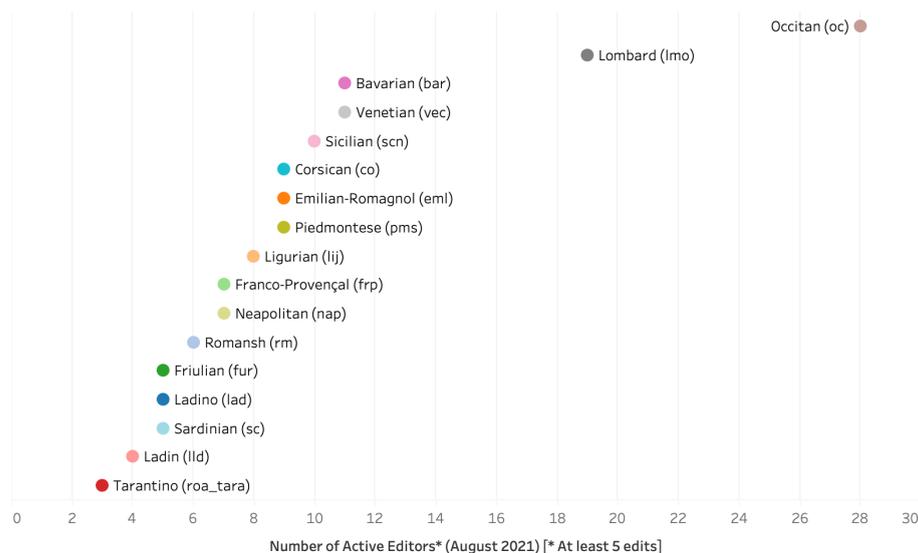


Wikipedia has reached its second decade, being the largest multilingual and collaborative free knowledge repository in human history. But, scientific studies over the past ten years have shown that Wikipedia has been unable to continue growing its editor communities. But is this decline of Wikipedia communities true? Generalizable?

In the following graph, we see the number of active editors of a group of Wikipedia language editions from Italy and the countries surrounding it. Despite the numbers varying across the languages, they all present a similar stagnation after peaking between 2008-2012. **Italian Wikipedia** is the fourth Wikipedia in the region in the number of active editors (3,259) in August 2021.



In the following graph, we zoom on the smallest Wikipedia language editions that were not visible in the previous one, and for one particular month. We see that most of them have less than 20 active editors in August 2021.



Are all communities in stagnation in the same situation? Are they at risk of dying out because of lack of renewal? We will zoom on the “active editor community” using a quantitative analysis approach to answer these questions.

Six indicators for community growth and renewal

We propose the creation of **6 indicators** that we call Vital Signs. In Medicine, vital signs indicate the status of the body's vital (life-sustaining) functions. These measurements are taken to help assess the general physical health of a person, give clues to possible diseases, and show progress toward recovery.

In the case of Wikipedia, Vital Signs are related to the community's capacity and function to grow or renew itself. Three of them are focused on the entire group of "active editors" creating content: **retention**, **stability**, and **balance**; the other four are related to more specific community functions: **admins**, **specialists**, and **global community participation**. We believe that obtaining values for the current "active community capacities" in these areas can constitute a reference point to plan to guarantee "openness" in these areas, and at the same time, to observe growth and renewal, and foresee and prevent future risks (e.g., bus factor).

Target values for healthy community renewal and growth

We think that a healthy community is in constant renewal with possibilities of growth. Based on our interpretation of the data for a wide array of Wikipedia language editions, we have estimated a set of target values for each of the Vital Signs to indicate renewal and growth.

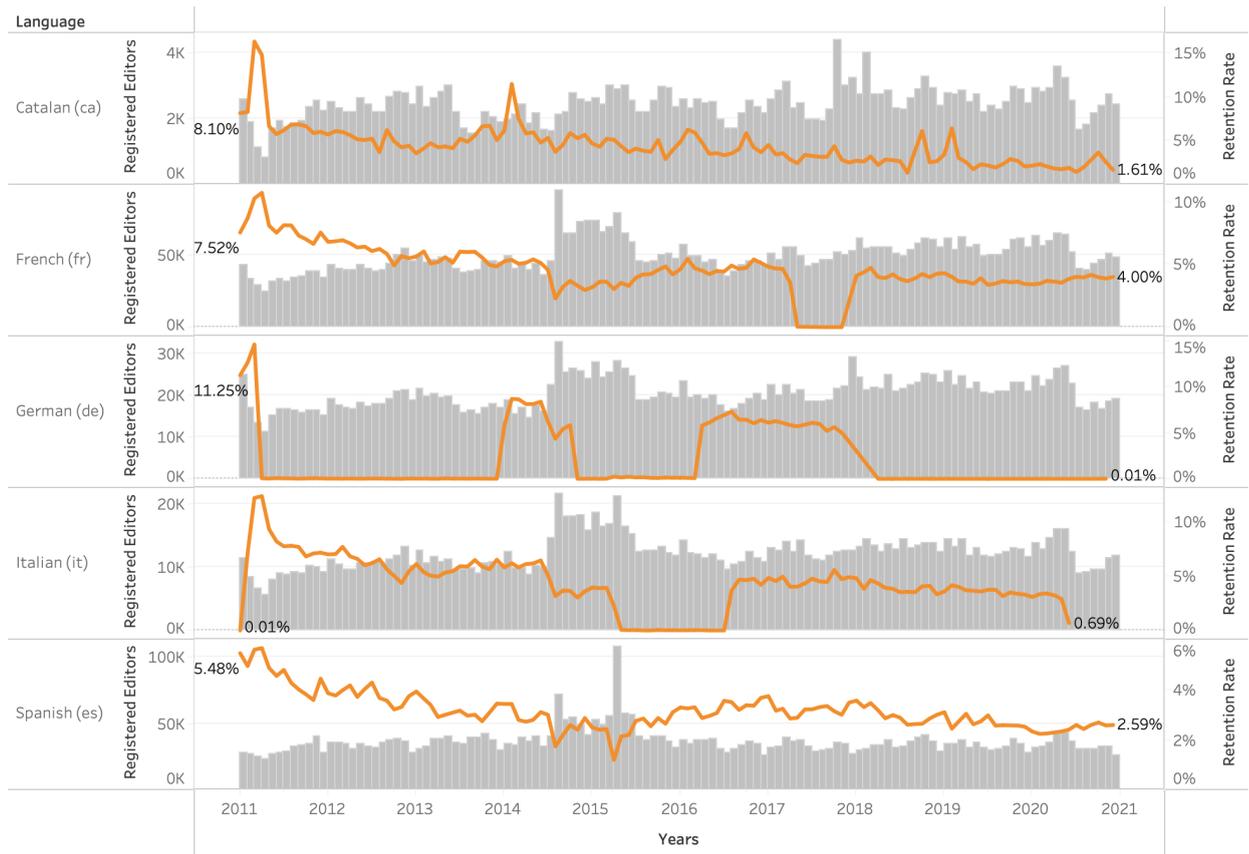
We think that these targets are reasonable for established communities. Smaller communities' goal is to get to a critical mass that allows them to be sustainable (minimum of 20 active editors per month).

1. Retention

The first vital is retention. This reflects the capacity of the community to engage new editors to continue editing after they register.

Metric: the retention rate is computed as the % of new editors who edit at least once 60 days after their first edit.

Description: This dual-axis graph shows us (left-axis) the number of registered editors by month (bars) and (right-axis) the retention rate as an orange line. We clearly see that the number of registered editors is stable or even increasing, while the retention rate is decreasing over time.





Target 1.- When Wikipedia language editions were growing, the retention rate was in the range between 25% and 5%. Therefore, a rate that is lower than this might not be sufficient to grow the community. We would set the target of a 3% retention rate to ensure there is renewal among editors but aim at 5-7%.

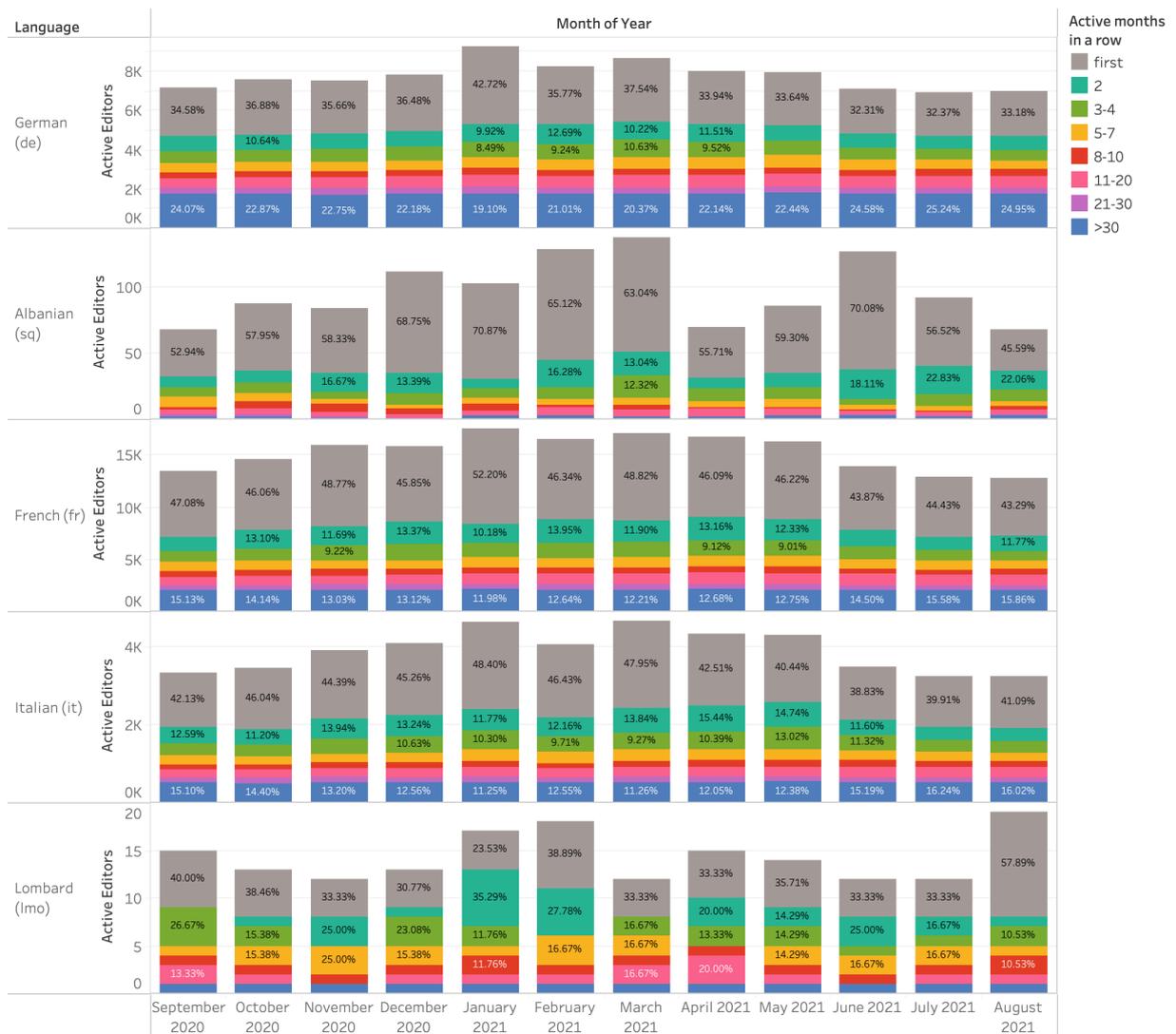
2. Stability

The second vital sign is stability. Community stability is the persistence of active editors. You do not only want to ensure that there are fresh editors every month who had not edited on the previous month, but also that there are others who have edited for many more months.

Metric: number of active editors by number of months they have been active in a row.

Description: In this graph, we see the number of active editors on a monthly basis, and in color the number of months in a row they have been editing for each month. Grey means the first month, which can either be the first month ever or the first month after a break of editing.

In Polish Wikipedia, fresh editors in a given month are about 35-40% (in grey), while in others like German, is around 33-35%. This means that Polish Wikipedia is less volatile than German. Polish is engaging fresh editors every month, but the active community is mostly composed of editors who continue one month after another.



Target 2.- A share of “fresh” editors of 30%-40% may be desirable to have an influx of new energy and ideas. If higher than this, and especially over 60%, it may be an indicator of a lack of capacity to

engage and stabilize the community. High percentages of fresh editors are only desirable when the number of active editors is growing.

Target 3.- A share of “long-term engaged” editors (including the bins of editing 13-24 months in a row and > 24 months in a row) around 33% is indicative of a solid community able to carry on with long-term Wikiprojects and activities. This seems a desirable target.

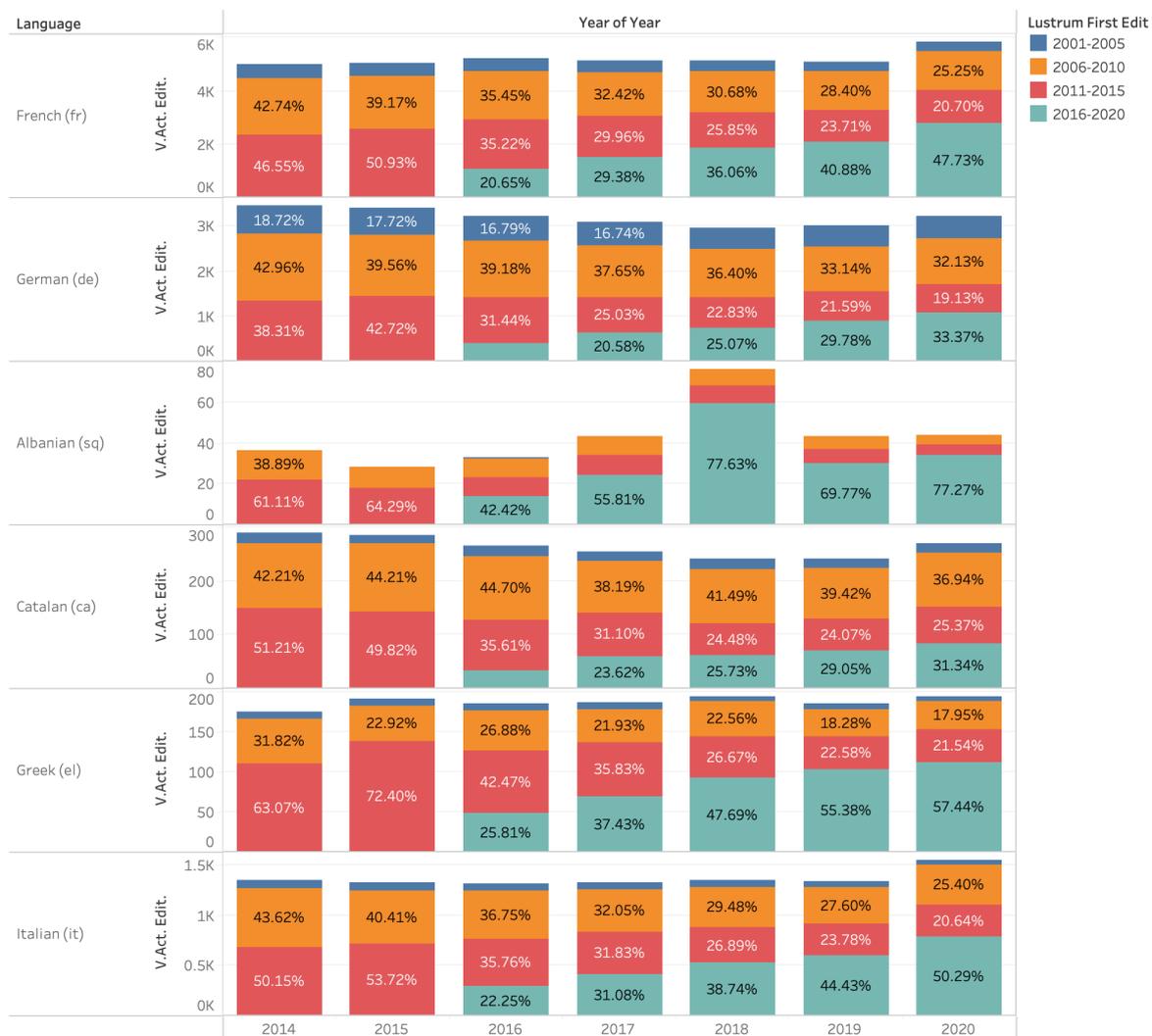
3. Balance

The third vital sign is balance. Community balance is being able to maintain an equitable proportion of old and new editors. We want to benefit from the experience, but also be able to stay open to new generations. This is a key sign indicating renewal. As it may not be desirable that the “productivity” relies too much on an older generation, but neither that it would depend mostly on the last one.

Metric: number and percentage of very active editors* by year and by generation (lustrum of the first edit). * at least 100 edits.

Description: In the following graph, we see for the selected language editions, the composition of the “very active editors” every year by lustrum of first edit (2001-2006, 2006-2010, 2011-2015, 2016-2020, and 2021-2025.) We could say that they are “generations”. By definition, very active editors are those who make at least 100 edits per month. Very active editors account for 80% of the edits made by humans every month. So, we could say that they are a group of very valuable editors. The graph shows the yearly number of editors who have been “very active” for at least one month.

We see that every year the percentage of editors who started editing during 2016-2020 is growing, mostly at expense of the 2006-2010 and 2011-2015 generations. The percentage of very active editors who started editing during 2001-2005 has not varied. In Polish Wikipedia, this is 9.13% in 2020, but 9.14% in 2019, 8.45% in 2018, and 10.18% in 2017.



Target 4.- A growing share of the last generation until occupying between 30 - 40% seems reasonable for a language edition that is not in a growth phase - larger when it is. We considered every generation to be 5 years (a lustrum), so, as a rule of thumb, 10-15% from the last generation should be incorporated every year into the community of very active editors.

Target 5.- A share of editors of at least 5-15% from the first generation (typically 2001-2005) seems a desirable target as well. Although they might be at the end of their lifecycle and the growth occurred with the following generation (2006-2010).

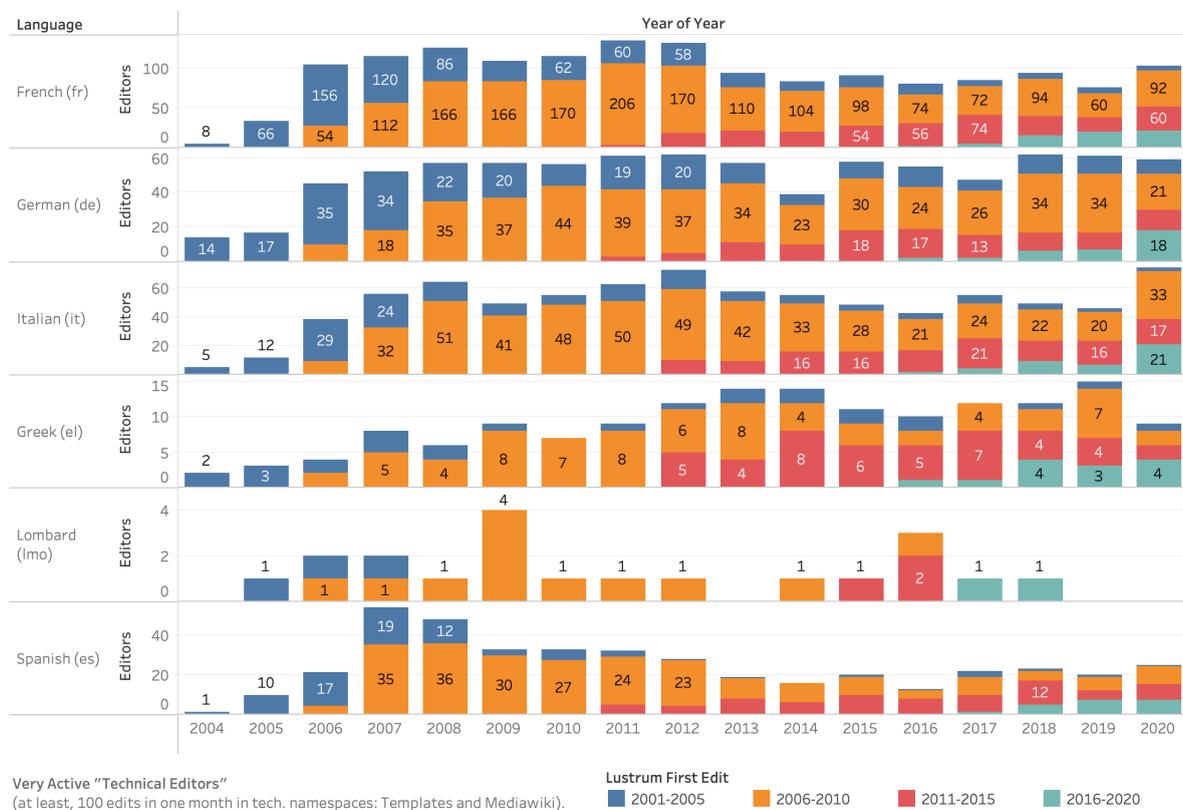
4. Special functions

The fourth vital sign is special functions. Community technical and coordination functions are undertaken by editors, are essential for the project.

4.1 Very Active Technical Editors

Metric: number of very active editors in technical namespaces* (at least one month) by year and by generation. * Mediawiki and Templates

Description: Similarly as with the previous measurement, here we focus on the very active technical editors. Editors who have done more than 100 edits in one month in technical namespaces (that is templates and MediaWiki namespaces). In the graph, we see the number of “Very active technical contributors”.



Target 6.- Having a minimum number of “very active technical editors” (at least 20) seems desirable considering the different tasks (bots, templates, etc.).

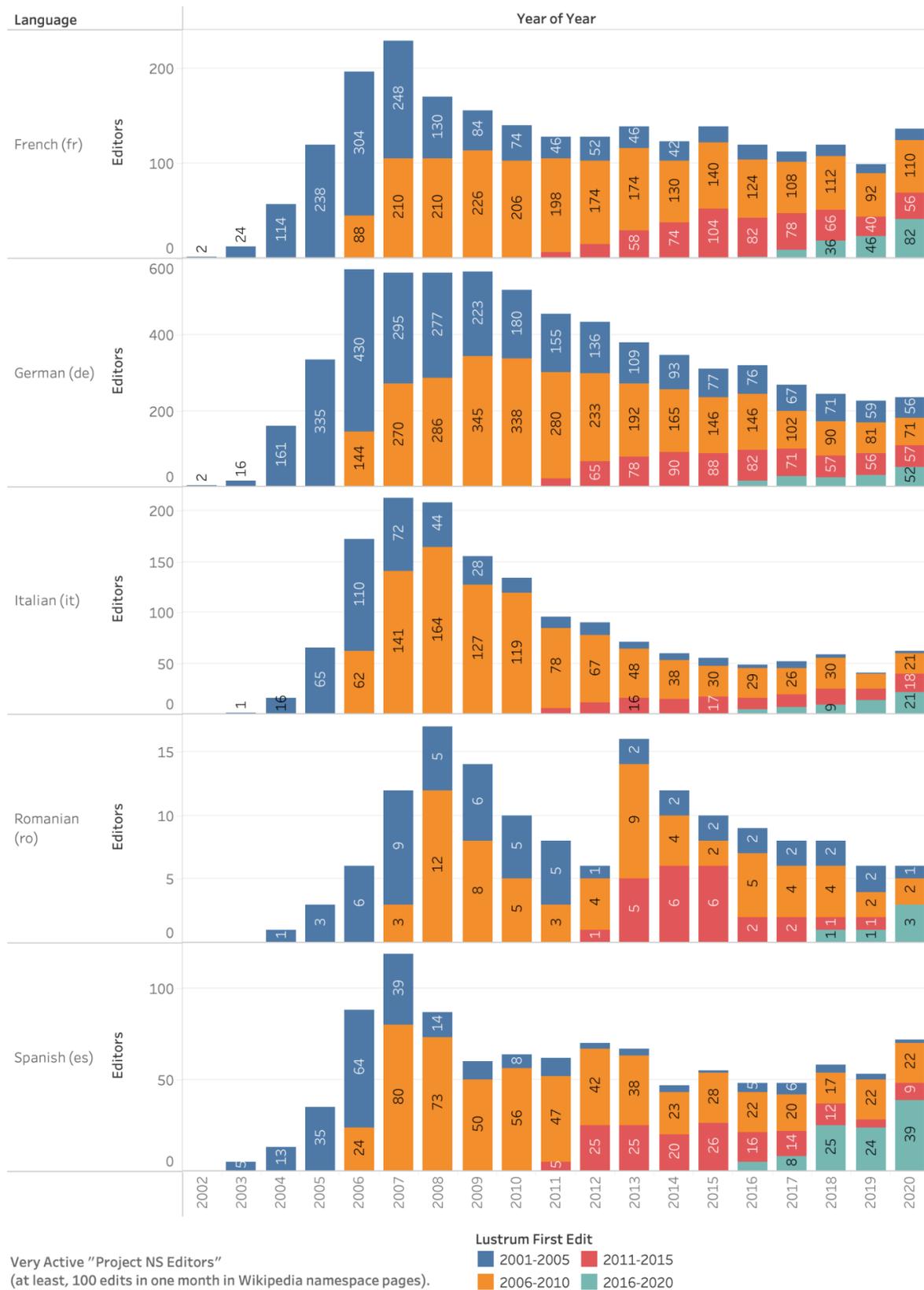
Target 7.- Given the usually low number of very active technical editors and the remarkable effort this role requires, it would be preferable that the majority were from newer generations. Renewal is key. A target of 30% from the last generation seems reasonable.

4.2 Very Active Coordinators

Metric: number of very active editors in coordination namespaces* (at least one month) by year and by generation

* Wikipedia and Help

Description: In this other case, we do exactly the same analysis but for those very active editors in the Help and Wikipedia namespace (this is, Wikiprojects, Village Pump, among others). The number of editors is higher. In the graph, we see the number of “Very active project coordinators”.



Target 8.- The number of coordinators (“very active editors” in Wikipedia namespace) should always be larger than the number of very active technical editors, since taking coordination activities is key to engaging editors into contests, Wikiprojects, among others.

Target 9.- The proportion of coordinators should be a minimum of 5-15% of the very active editors to guarantee there are common initiatives that involve everyone. A low proportion of coordinators implies that very active editors are working in silos.

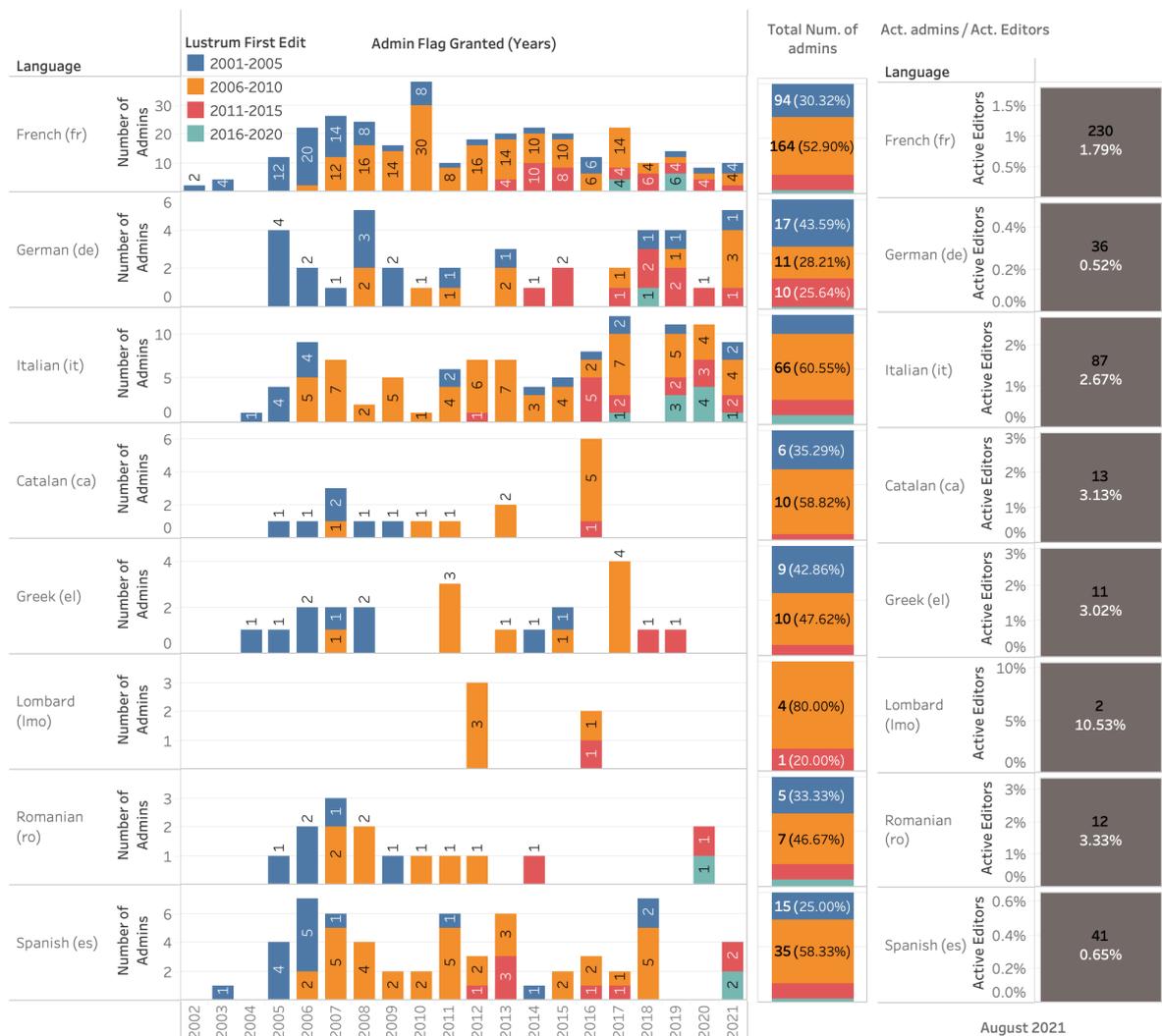
5. Administrators

The fifth vital sign is admins. Admins have rights and responsibilities in performing actions over the content and take a key function for the community.

Metric: number of admins by year of flag granted and by generation; total number of admins by generation; ratio between the number of active admins and the number of active editors.

Description: On the left subgraph, we see the admins' flags granted by year, and the color represents the generation they belong to. Most flags were granted from 2011 to 2016 and to previous generations. We almost do not see flags from editors 2016-2020, but Italian (8, which are a percentage below 10%). No flags have been granted in the Catalan Wikipedia after 2015.

In the middle graph, we see the total number of admins by generation. On the very right, we see the number of active admins in the past August 2021. The percentage is the ratio between the number of active admins by the number of active editors. This percentage varies according to the language, but it tells the "load" each admin is carrying, given that their task is to patrol the production and act when necessary. The lower the percentage, the higher the load they take.



Target 10.- To guarantee openness to positions of responsibility and privilege, there should be new admins every year (e.g., at least 5% of the total number of admins). This may imply setting an expiry date for the role or a voluntary request to lower-activity admins to renounce the role.

Target 11.- The overall group of admins should be balanced in terms of the different generations in which they started editing.

Target 12.- The proportion of admins among active editors should be from 1% to 5% to guarantee that admins do not carry an excessive workload, since, in the end, they revise other editors' edits.

Target 13.- Any community should have a minimum of 3 admins, regardless of their size in active editors. Even though the work may be low, a minimum number of admins guarantees the continuity of work.

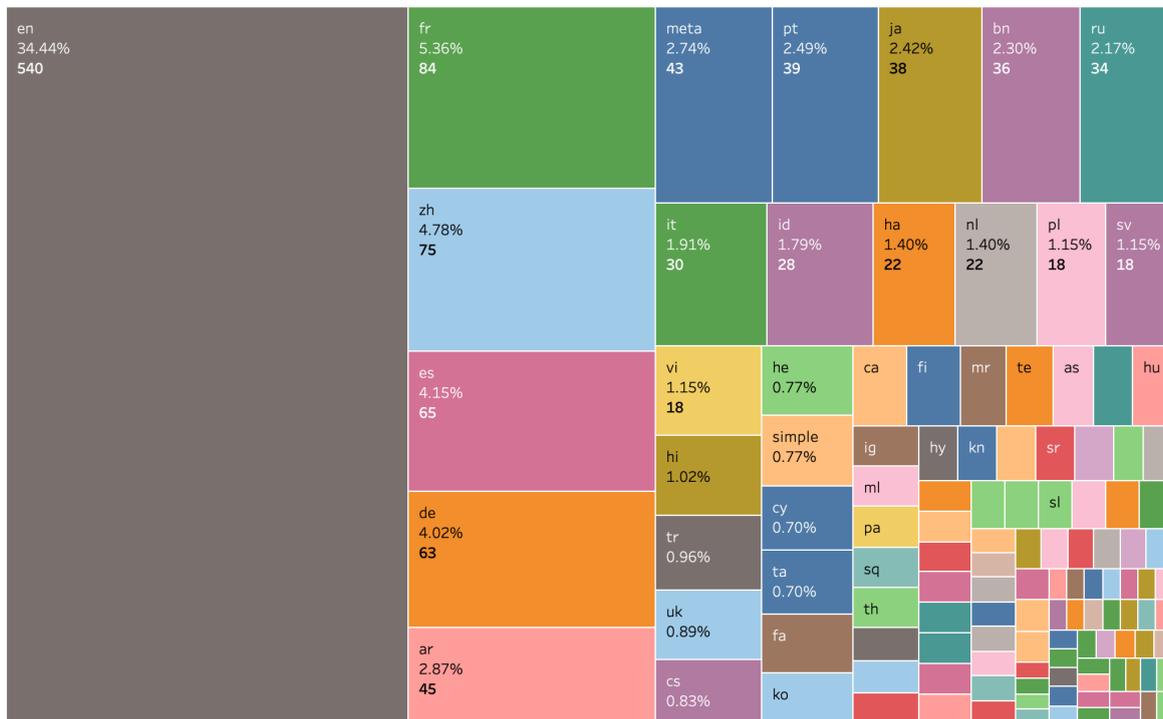
6. Global

The sixth and last vital sign is global. Communities' participating in the "global community" is key to making their voice heard and learning from others.

Metric: ration between the number of active editors in Meta-wiki from a Wikipedia language edition and the number of active editors in that Wikipedia language edition.

Description: In this graph, we see the number of active editors in a Wikipedia language edition (top) and the number of active editors in Meta-Wiki whose primary language edition is the language edition (bottom). We can see that the ratio between the two is around 0-2% (e.g., 1.91% of Italian WP in August 2021).

Meta-Wiki Active Editors by Primary Language (August, 2021)

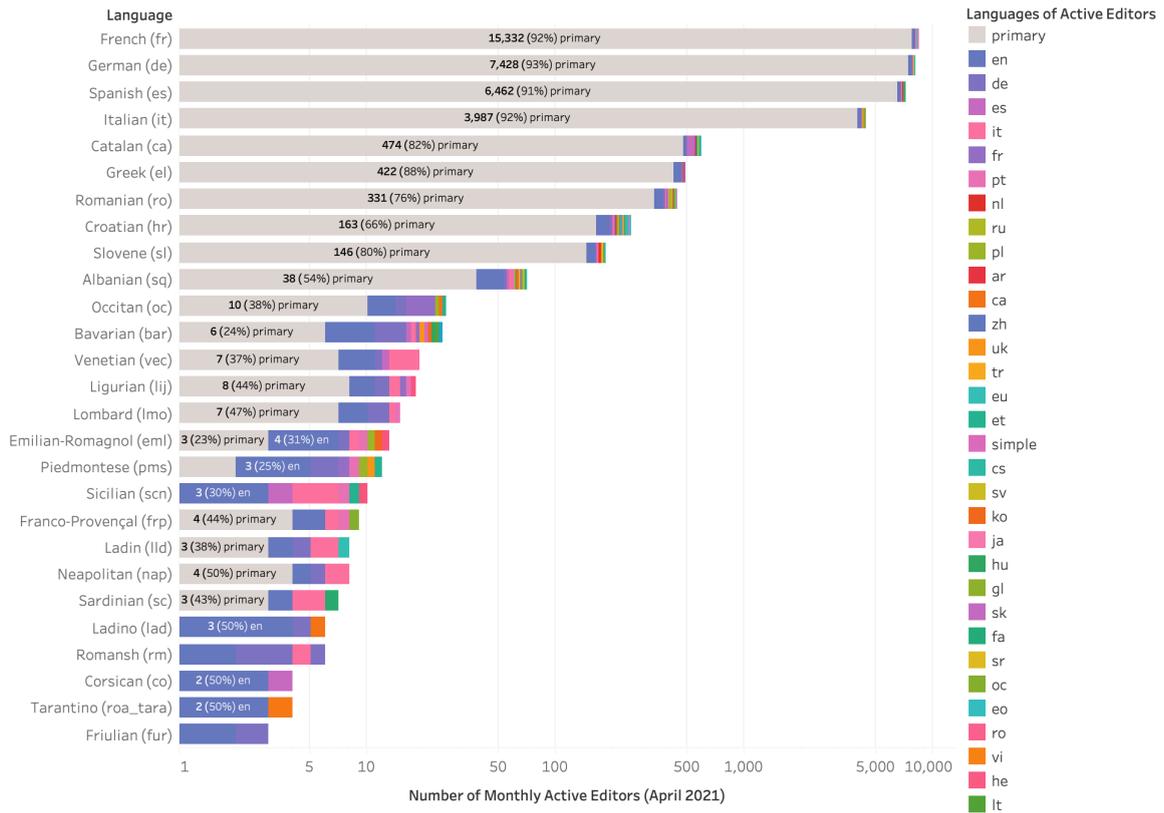


Target 14.- The number of editors from a Wikipedia language edition community active in Meta-wiki should be around 1% of the active editors. A much lower value would imply that the language edition community is not participating enough in the global movement.

Metric: number and percentage of active editors in a Wikipedia language edition by primary language. An editor's primary language edition is the one in which they made more edits.

Description: In this graph, we can see the number of active editors in Italian area (Italy and neighbour countries) Wikipedia language editions in August 2021, and the percentage of editors from the same language (primary).

While Italian and German have very high percentages of primary editors, other languages like Albanian, or the regional languages are below 60%. Even though the non-primary editors surely make some useful edits, one could say that in terms of developing Wikiprojects and taking responsibility, the real community tends to be the one composed by primary editors.



Target 15.- The proportion of primary language editors among the active editors in a language community should be at least 60% to guarantee that there are dedicated editors whose main project is that Wikipedia language edition. Higher than 90% might imply that the community is not attracting “sporadic collaborators”.